

CLAIMS:

1. An electronic data discovery apparatus, comprising:
 - a data access and merge device configured to download a plurality of files to be filtered from at least one archive;
 - a de-duplication device configured to compare files from at least one custodian and tag one of said files from at least one custodian with duplication information; and
 - a criteria filtering device configured to screen said plurality of files against at least one of a compliance word and a privilege word.
2. The apparatus of Claim 1, said data access and merge device comprising:
 - an predetermined data structure derived tagging module configured to tag a file with a unit of file meta-data.
3. The apparatus of Claim 2, the unit of file meta-data comprising at least one of:
 - file name;
 - date last modified;
 - date created;
 - author; and
 - subject.
4. The apparatus of Claim 1, said data access and merge device further comprising at least one of:
 - a virus identification and cleaning device;
 - an encryption/password identification and decryption/key recovery device;and
 - a foreign language identification and conversion device.
5. The apparatus of Claim 1, said de-duplication device comprising at least one of:

a vertical de-duplication device configured to compare files from a single custodian and tag one of said files from a single custodian with vertical duplication information; and

a horizontal de-duplication device configured to compare files from a plurality of custodians and tag one of said files from a plurality of custodians with horizontal duplication information.

6. The apparatus of Claim 5, said vertical de-duplication device comprising:

- a meta-data comparison device;
- a content comparison device;
- a file binary comparison device; and
- a time stamp comparison device.

7. The apparatus of Claim 5, said horizontal de-duplication device comprising:

- an author/originator filtering device;
- a meta-data comparison device;
- a content comparison device;
- a file binary comparison device; and
- a time stamp comparison device.

8. The apparatus of Claim 1, said criteria filtering device comprising:

- a compliance word filtering device configured to screen said plurality of files against a predetermined compliance word so and produce one of a compliant file and a non-compliant file;

- a privileged word filtering device configured to screen said compliant file against a predetermined privileged word and produce one of a compliant, privileged file and a compliant, non-privileged file;

- a production set storage device configured to store said compliant, non-privileged file; and

- a privileged set storage device configured to store said compliant, privileged file.

9. The apparatus of Claim 8, further comprising:
an index scheme selection device configured to store at least one of said predetermined compliance word and said predetermined privileged word; and
a synonym set creation device configured to store a synonym at least one of said predetermined compliance word and said predetermined privileged word.

10. The apparatus of Claim 8, further comprising:
a file converter device configured to convert one of said compliant, non-privileged file and said compliant, privileged file to a production file; and
a profiler configured to estimate at least one of a printed page count and a cost to print said production file.

11. A system for electronic data discovery, comprising:
an electronic data discovery apparatus configured to produce at least one of a compliant, non-privileged file and a compliant, privileged file; and
a production device configured to produce a production file corresponding to said at least one of a compliant, non-privileged file and a compliant, privileged file, said electronic data discovery apparatus comprising
a data access and merge device configured to download a plurality of files to be filtered from at least one archive,
a de-duplication device configured to compare files from at least one custodian and tag one of said files from at least one custodian with duplication information, and
a criteria filtering device configured to screen said plurality of files against at least one of a compliance word and a privilege word.

12. The system of Claim 11, said data access and merge device comprising:
an predetermined data structure derived tagging module configured to tag a file with a unit of file meta-data.

13. The system of Claim 12, the unit of file meta-data comprising at least one of:
file name;

date last modified;
date created;
author; and
subject.

14. The system of Claim 11, said data access and merge device further comprising at least one of:

a virus identification and cleaning device;
an encryption/password identification and decryption/key recovery device;
and
a foreign language identification and conversion device.

15. The system of Claim 11, said de-duplication device comprising at least one of:

a vertical de-duplication device configured to compare files from a single custodian and tag one of said files from a single custodian with vertical duplication information; and

a horizontal de-duplication device configured to compare files from a plurality of custodians and tag one of said files from a plurality of custodians with horizontal duplication information.

16. The system of Claim 15, said vertical de-duplication device comprising:

a meta-data comparison device;
a content comparison device;
a file binary comparison device; and
a time stamp comparison device.

17. The system of Claim 15, said horizontal de-duplication device comprising:

an author/originator filtering device;
a meta-data comparison device;
a content comparison device;
a file binary comparison device; and

a time stamp comparison device.

18. The system of Claim 11, said criteria filtering device comprising:
a compliance word filtering device configured to screen said plurality of files against a predetermined compliance word so and produce one of a compliant file and a non-compliant file;

a privileged word filtering device configured to screen said compliant file against a predetermined privileged word and produce said one of a compliant, privileged file and a compliant, non-privileged file;

a production set storage device configured to store said compliant, non-privileged file; and

a privileged set storage device configured to store said compliant, privileged file.

19. The system of Claim 11, said electronic data discovery apparatus further comprising:

a data export device configured to export at least a portion of a file to a remote processor via one of a network connection, direct connection, a wireless connection, and a portable media drive, wherein

said remote processor is configured to perform at least one of store a filter result, remove a virus, and decrypt/unprotect a file.

20. The system of Claim 11, said electronic data discovery apparatus further comprising:

an external control device configured to receive instructions and provide status information to a remote control device.

21. The system of Claim 11, said electronic data discovery apparatus further comprising:

a filter results storage device configured to store results of a filter operation;
and

a printer connection device configured to relay information to a printer.

22. A method for performing electronic data discovery, comprising:

downloading a plurality of files to be filtered from at least one archive;
comparing files from at least one custodian and tagging one of said files from at least one custodian with duplication information; and
screening said plurality of files against at least one of a compliance word and a privilege word.

23. The method of Claim 22, said downloading a plurality of files comprising:
tagging a file with a unit of file meta-data derived from an predetermined data structure.

24. The method of Claim 23, the predetermined data structure comprising at least one of:
file name;
date last modified;
date created;
author; and
subject.

25. The method of Claim 22, said downloading comprising at least one of:
identifying and cleaning a virus;
identifying an encrypted/password-protected file and decrypting/unprotecting said encrypted/password-protected file; and
translating a foreign language file to a predetermined language.

26. The method of Claim 22, said comparing files comprising at least one of:
vertical de-duplicating files from a single custodian and tagging one of said files from a single custodian with vertical duplication information; and
horizontal de-duplicating files from a plurality of custodians tagging one of said files from a plurality of custodians with horizontal duplication information.

27. The method of Claim 26, said vertical de-duplicating comprising:
comparing file meta-data;

comparing file contents
comparing file binaries; and
comparing file time stamps.

28. The method of Claim 26, said horizontal de-duplicating comprising:
comparing file author/originator information;
comparing file meta-data;
comparing file contents
comparing file binaries; and
comparing file time stamps.

29. The method of Claim 22, said screening comprising:
compliance word filtering said plurality of files against a predetermined
compliance word and producing one of a compliant file and a non-compliant file;
privileged word filtering said compliant file against a predetermined
privileged word and producing one of a compliant, privileged file and a compliant,
non-privileged file;
storing said compliant, non-privileged file in a production set storage device;
and
storing said compliant, privileged file in a privileged set storage device.

30. The method of Claim 29, further comprising:
storing at least one of said predetermined compliance word and said
predetermined privileged word in an index scheme selection device; and
storing a synonym at least one of said predetermined compliance word and
said predetermined privileged word in a synonym set creation device.

31. The method of Claim 29, further comprising:
converting one of said compliant, non-privileged file and said compliant,
privileged file to a production file; and
estimating at least one of a printed page count and a cost to print said
production file.

32. An apparatus for electronic data discovery, comprising:

means for downloading a plurality of files to be filtered from at least one archive;

means for comparing files from at least one custodian and tagging one of said files from at least one custodian with duplication information; and

means for screening said plurality of files against at least one of a compliance word and a privilege word.

33. A computer program product storing instructions for execution on a computer system, which when executed by the computer system, causes the computer system to perform the method recited in any one of Claims 22-31.